



Published in final edited form as:

Int J Geriatr Psychiatry. 2019 April ; 34(4): 578–587. doi:10.1002/gps.5054.

Verbal Fluency in a National Sample: Telephone Administration Methods

Janice C. Marceaux^a, Michelle A. Prosjje^b, Leslie A. McClure^c, Bhumika Kana^d, Michael Crowe^e, Brett Kissela^f, Jennifer Manly^g, George Howard^h, Joyce W. Tamⁱ, Frederick W. Unverzagtⁱ, and Virginia G. Wadley^d

^aDepartment of Veterans Affairs, South Texas Veterans Health Care System, San Antonio, TX

^bNeuroBehavioral Specialists of Jacksonville, Jacksonville, FL

^cDepartment of Epidemiology and Biostatistics, Drexel University, Philadelphia, PA

^dDepartment of Medicine, University of Alabama at Birmingham, Birmingham, AL

^eDepartment of Psychology, University of Alabama at Birmingham, Birmingham, AL

^fDepartment of Neurology, University of Cincinnati, Cincinnati, OH

^gDepartment of Neurology, Columbia University, New York, NY

^hDepartment of Biostatistics, University of Alabama at Birmingham

ⁱDepartment of Psychiatry, Indiana University School of Medicine, Indianapolis, IN

Abstract

Objectives: Describe novel methods for ascertaining verbal fluency in a large national sample of adults, examine demographic factors influencing performance, and compare scores to studies using in-person assessment.

Methods/Design: Participants were from the REasons for Geographic and Racial Differences in Stroke (REGARDS) study, a national, population-based, longitudinal study of stroke in adults aged 45 years and older. Letter and semantic fluency were gathered, using Letter “F” and Animal Naming, via a telephone-based assessment with computer-assisted scoring of digital recordings.

Results: Initial letter and semantic fluency scores were obtained on 18,505 and 18,072 participants, respectively. For both fluency tests, scores were normally distributed. Younger age and more years of education were associated with better performances ($p < 0.0001$). The mean and standard deviation for matched subgroups, based on age, gender, and education, were quite

Corresponding author: Janice C. Marceaux, PhD, ABPP, Psychology Service (116B), South Texas Veterans Healthcare System, 7400 Merton Minter Blvd., San Antonio, TX 78229. **Alternate corresponding author:** Virginia G. Wadley, PhD, Department of Medicine, Division of Gerontology, Geriatrics & Palliative Care, University of Alabama at Birmingham, 1720 2nd Avenue So., CH19-218T, Birmingham, AL 35294.

Footnote: The cohort was recruited from the 48 contiguous United States using telephone and mailing methods followed by home visits for blood, urine, ECG, and anthropometric measures conducted by Examination Management Services, Inc. All recruitment and follow-up calls, including cognitive assessments, were centrally administered by staff of the Survey Research Unit in the School of Public Health at the University of Alabama at Birmingham (UAB). ECGs were read at Wake Forest University, blood is stored at the University of Vermont, fluency assessments were scored at Indiana University and UAB, and analyses were conducted at UAB and Drexel University.

comparable to scores reported out of samples using an in-person administration format. Telephone-based assessment also allowed for a level of quality control not available via in-person measurement.

Conclusions: Telephone-based assessment of verbal fluency and computer-assisted scoring programs designed for this study facilitated large scale data acquisition, storage, and scoring of protocols. The resulting scores have similar characteristics to those obtained by traditional methods. These findings extend validation of cognitive assessment methods, using survey research staff and computer-assisted technology for test administration.

Keywords

neuropsychological test; verbal fluency; telephone-based assessment; computer-assisted scoring; cognitive assessment; validation

Introduction

Telephone-based cognitive testing allows for the efficient assessment of large numbers of geographically dispersed persons as is often required in epidemiological and survey research. One approach is to modify an existing test for telephone use by adding or deleting items (e.g., the Telephone Interview for Cognitive Status), and another is to administer an existing in-person test unchanged over the telephone (e.g., verbal fluency). Two recent reviews, reporting on 75 studies, identified 20 telephone-based cognitive tests--a mix of modified and existing cognitive measures--that have been used to characterize cognitive status in adults in epidemiological and survey studies^{1,2}. The reviews identified important characteristics of these tests, including item content, administration time, and relation to clinical status, but also limitations including a lack of direct comparisons to other cohorts¹.

In this paper, we describe a telephone-based assessment of verbal fluency, an unmodified in-person measure of cognition, in an epidemiological context (initial cohort of 30,239 participants). We outline the approach to training and monitoring of data gatherers and scorers, describe a computer-based interface that assists in administration and scoring, report the error rates for this method of assessment, and compare score distributions for age- and education-matched subgroups to previously published studies that used in-person assessment.

Methods

Sampling Frame and Study Design

The REasons for Geographic And Racial Differences in Stroke (REGARDS) study is an ongoing national, population-based, longitudinal study of racial and geographic disparities in stroke, stroke mortality, and cognitive decline³. From commercially available lists, community-dwelling adults were randomly selected. Those aged 45 years or older and either African American or white were eligible for the study. Exclusion criteria included self-reported medical conditions (e.g., cancer) that would prevent long-term participation, being on a waiting list for nursing home, or being unable to participate in telephone interviews due to the interviewer's judgment of poor cognition or hearing deficits. The sample was

recruited between January 2003 and December 2007 using mail, telephone, and in-home contacts (33% response rate, 49% cooperation rate)⁴. The cohort consists of 30,239 individuals, with oversampling of residents in stroke belt states (56%; NC, SC, GA, AL, MS, TN, AR, and LA) and the remainder residing in the other 40 contiguous states in the U.S. The cohort is composed of 45% men, and 42% African American participants.

Cognitive tests selected for REGARDS were short, amenable to telephone administration, and sensitive to cognitive changes associated with stroke and vascular risk factors^{5–7}. Semantic and letter fluency tests, which are the focus of the present investigation, have high test-retest reliability (>.70 across studies) and high inter-rater reliability (>.97 across studies), with moderate to high correlations among letter (ranging from .72 - .94) and semantic category variants (.66 - .71)⁸.

The analytic sample included participants who were stroke-free (by self-report) at baseline and through the time of the first verbal fluency assessment (by expert review of medical records) and whose electronic sound files were scored by April 1, 2016. Study procedures were reviewed and approved by the institutional review boards at the collaborating institutions. All participants provided informed consent to participate.

Measures

Animal Naming was first implemented in January 2006 with follow-up assessments every 24 months. For this analysis, only initial assessments are included. After standard instructions and examples, participants attempt to generate as many animal names as possible in 60 seconds. Letter Fluency, using letter “F”, began in February 2008 with follow-up assessments every 24 months. Again, only initial assessments are included. After standard instructions and examples, participants attempt to generate as many words beginning with the letter “F” as possible in 60 seconds.

Age, education (years completed), race (African American or white), sex (male, female) and health history were assessed by self-report. An in-home examination was used to gather physiological measures (e.g., blood pressure, blood/urine samples, electrocardiogram) and an inventory of current medications. Incident stroke was ascertained via telephone follow-up every six months using the Questionnaire for Verifying Stroke-free Status (QVSS)⁹ and verified by medical record review and adjudication by a panel of neurologist stroke experts. Specific definitions and classification criteria for vascular risk factors (e.g., diabetes, heart disease, atrial fibrillation) has been described elsewhere¹⁰.

Procedures

Computer Assisted Assessment—Administration scripts were programmed into a computer-assisted telephone interview (CATI). Participants gave verbal permission to record responses onto electronic sound files stored on a dedicated secure server using WinCati 4.2/ Sawtooth Ci3. Interviewers read administration instruction scripts from a computer screen. Once a participant completed the examples properly, the interviewer pressed a key to begin an audio recording. Scores consist of the total number of valid responses for each participant, the valid number in each 15 second epoch, and intrusions and repetitions. For

Animal Naming, valid responses included any living or extinct animal, bird, or insect. Credit was given for superordinate categories (e.g., fish) as well as subordinate exemplars (e.g., trout). Gender distinctions (e.g., mare, stallion) and age distinctions (e.g., calf, cow) were given credit. Repetitions were defined as exact repetitions and singular and plural forms of the same animal name (e.g., dog, dogs). Intrusions consisted of any non-animal words. For Letter Fluency, intrusions were defined as a response beginning with a letter other than the letter 'F' to which the interviewer provided the following prompt after the first such error, "No, say words that start with the letter 'F' as in 'Fred'." This prompt was also used if the participant began by generating words beginning with the "ph" phoneme rather than the letter "F". Scores consist of the total number of valid responses including potential homophones (e.g., feet and feat, fair and fare, and fax and facts). Repetitions were defined as exact repetitions as well as variations in tense or number, or providing the same root word with different endings (e.g., "finds" or "finding" after having given "find"). Intrusions consisted of any words beginning with a letter other than "F" and any non-existent words.

Computer-Assisted Scoring—A computer-assisted scoring interface was used to review and score files (see Supplemental Information). The scoring interface for each includes the following basic functions: 1) start/stop buttons, 2) counter button to tally the number of responses, 3) boxes for scorers to enter the numbers of repetitions and intrusions, and 4) buttons to save, view, or calculate the scores. Additional features include a timer to indicate the length of a recording, a pause button to allow pausing at desired intervals during review, and a system of double data scoring for reliability analyses. This information is populated into the box on the left side of the screen, which also displays participant ID and sequence of administration, scores (V1=initial score; V2=second score), and flags for administration errors.

Several categories of administration violation were captured by scorers for ongoing quality control and retraining of interviewers. These include: *Prompt* ("unscripted" dialogue or assistance by an interviewer, e.g., "think of the zoo" during Animal Naming), *Help* (someone in the participant's home is heard offering assistance in responding to the task), *Invalid* (interviewer stopping the recording too early or equipment failure during data collection), and *No Audio* (no recording captured). Scorers clicked buttons to assign responses to the overall total and to the appropriate 15-second interval and to flag responses as repetitions or intrusions. The scoring program allows data to be linked to each particular interviewer and computer station, which facilitates targeted quality control.

Interviewer and Scorer Training—A total of 95 interviewers were trained by the director of the Survey Research Unit using a checklist including all elements of proper administration (scripted instructions, starting/stopping the 60 second recording accurately, remaining attentive, and providing allowed prompts when indicated). In a classroom-type setting, 160 hours of training were conducted for cognitive assessments including the fluency measures, which took place over 20 sessions (5 days a week, 8 hours a day, 4 weeks). Instruction included lecture, demonstration, and experiential exercises to acquaint trainees with the CATI interface after which trainees made "practice calls" to supervisors and other interviewers in an effort to simulate actual studies. After two consecutive practice

calls were administered without error, the interviewer was certified to administer the assessments. Supervisors then routinely monitored each certified interviewer a minimum of four times per month by scrolling through the interview screens and listening in to randomly selected “live” interviews from a separate supervisory CATI station. Thus, ongoing correction and retraining could be implemented as needed. Annual 90-minute refresher training for the administration and recording of the cognitive battery was done for all certified interviewers. These sessions focused on the rationale for cognitive assessments, importance of standardized procedures, and review of common administration errors and challenging participant behaviors.

Scorers were trained in sessions led by experienced psychologists (Drs. Wadley and Kana). Ten scorers have been used to date; all were college graduates or current college students. Scorers first reviewed manualized instructions that included a list of atypical responses and situations. A set of 50 Animal Naming and 50 Letter Fluency sound files were used for certification. Each scorer had to achieve .88 kappa adjusted agreement or better relative to the scores assigned by Drs. Wadley or Kana. After certification, a randomly selected 10% of new files from each scorer were independently scored by Dr. Kana for ongoing monitoring of quality control. Any coefficients below .88 would result in suspension of a scorer’s scoring activities and a period of re-training.

Statistical Analyses

Of 30,239 REGARDS participants, 22,250 had first Letter Fluency scores by April 1, 2016. We excluded 1,104 participants due to self-reported stroke at baseline and 498 due to stroke after baseline but prior to cognitive assessment. Another 518 (2.8%) had missing data, 421 had invalid data, 35 were missing due to mismatched ID numbers, and 1,169 had no audio. Thus 18,505 participants remained for Letter Fluency analysis. For Animal Naming, there were 25,478 participants with an initial assessment. We excluded 1,408 with self-reported stroke at baseline, and 379 with stroke prior to first assessment. Another 140 had missing data, 5,167 had no audio, and 313 had invalid data. Thus, 18,072 participants remained for Animal Naming analysis.

Total score, repetitions, and intrusions were tallied by test type and participant subgroups and analyzed using Chi-square tests, t-tests, or analyses of variance (ANOVA) as appropriate. Age was categorized into five groups: 45–54; 55–64; 65–74; 75–84; 85+. Years of educational attainment was categorized into six groups: 0–8; 9–11; 12; 13–15; 16; and 17+.

Mean fluency scores and standard deviations of matched subsets of the REGARDS sample were also compared to scores obtained in published studies using in-person assessment. Specifically, using demographic characteristics reported in published studies, we used for comparison the mean score values of all REGARDS participants who matched those sample characteristics with respect to age range, race and gender composition, and education range.

Results

Among those who were invited to be recorded ($n=24,227$), 716 (3%) refused. Compared to those who agreed to having their responses recorded, those who refused were more likely to be older (67 vs. 65 yrs., $p=0.0001$), African American (4% vs. 2%, $p<0.0001$), men (4% vs. 3%, $p<0.0001$), and less educated ($p<0.0001$). There were no differences between those who did and did not agree to recording as a function of residence in the stroke belt vs. the non-belt regions, nor were there differences in the frequency of health conditions (e.g., diabetes, hypertension).

Sample characteristics by fluency test are presented Table 1. A total of 18,072 participants completed Animal Naming (56% female; 61% white; mean age 67.8 years [$SD=9.2$, range 46.0 to 99.0 years]; and 10% <high school [HS] graduate, 25% HS graduate, and 64% >HS graduate). A total of 18,505 participants completed Letter Fluency (57% female; 63% white; mean age 69.1 years [$SD=9.1$, range 46.5 to 99 years], and 9% <HS graduate, 25% HS graduate, and 66% >HS graduate). Chronic health conditions including hypertension and diabetes are at rates typical for an older adult community-dwelling sample, comparable to those reported by the National Health and Nutrition Examination Survey^{11,12}.

Score Distributions and Demographic Influences

Frequency distributions for each measure are presented in Figure 1 (panels A and B). Both measures are normally distributed. Table 2 shows mean, standard deviation, and range as a function of education, age and gender subgroups. As can be seen, for both semantic and letter fluency better performances are associated with younger age and more years of schooling completed (all p -values <0.0001). There are trivial, although statistically significant ($p<0.0001$), differences attributable to gender. Intrusion errors were rare, with less than 2% of the sample committing one or more intrusions on Animal Naming and approximately 6% of the sample committing one or more intrusions on Letter Fluency. Approximately 97% of participants had three or fewer repetitions on either test, with 60% having no repetitions for Animal Naming and 64% having no repetitions on Letter Fluency.

Comparison of Scores to Published In-Person Verbal Fluency Assessments

Table 3 presents mean scores and standard deviations from published studies including control group or normative sample verbal fluency performance obtained under standard, in-person administration procedures^{13–22} and compares those to demographically-matched subsamples extracted from the REGARDS study in which verbal fluency was measured via telephone-based procedures. In most cases, the Animal Naming mean and standard deviation for any given education, age, or gender subgroups obtained via standard methods are within 1–2 raw score points of the corresponding REGARDS telephone-based assessment method scores. The ranges are larger for Letter Fluency, with REGARDS matched samples tending to produce scores 2–6 raw score points lower.

Administration Violations

Violations of standard protocol during test administration included interviewer errors (coded as containing inappropriate *Prompt* or as *Invalid*); the *Invalid* files were excluded from

further analyses based on a priori decision that files less than 50 seconds in duration were not usable). Other violations involved assistance from a person in the participant's home (*Help* code). Files with technical difficulties (equipment failure, inaudible or *No Audio* code) could not be scored. Table 4 shows the prevalence of these violations for each type of fluency assessment. Technical difficulties affected 20% of the initial files for Animal Naming, which was introduced in 2006, and 5% for Letter Fluency, which was introduced in 2008 after improvements were made to the administration platform. Rates of disallowed prompting were low for both tests (<3%). The overall rate of assistance from a person who could be heard offering help in the participant's environment was <1%. Chi-square tests of association and t-tests or ANOVA were utilized to further examine the relationship between the occurrence of assists and participants' demographic characteristics. Education, gender, and age were significantly related to receiving assistance during administration of both Animal Naming and Letter Fluency, while race and region of residence had no association with receiving assistance. Specifically, those with lower educational attainment were more likely to have received assistance, as well as those who were older and/or male.

Discussion

We present details on the methodology and process for an individually administered, telephone-based, computer-assisted administration and scoring of verbal fluency used in the REGARDS study. Telephone-based assessment is increasingly important for epidemiological and survey research owing to its convenience, efficiency, and amenability to centralized quality control monitoring.

In our experience, telephone-based, computer-assisted administration of verbal fluency tests can be completed reliably and validly by typical survey workers. A smaller subset of individuals with college preparation and special training were able to score protocols in a detailed way with a high level of accuracy. The computer-assisted programs designed specifically for REGARDS facilitated large scale assessment and identification of several valuable scores related to performance and quality assurance. These programs also allow for further examination of cognitive processes in the structure of semantic knowledge through examination of response order, clustering, and switching.

Analyses indicated that verbal fluency assessed via telephone produced normal score distributions, as has been reported for in-person administration of the same measures. Analyses also indicated that age and education have the same relationship to verbal fluency when assessed over the telephone as they do in-person, that is, younger persons and those with more years of schooling tend to produce better performances. In a series of subanalyses, we demonstrated that REGARDS subgroups demographically-matched to published studies' control/normative samples yielded very close correspondence in mean scores and standard deviations across multiple authors and samples for the Animal Naming test. When deviations occurred, they tended to be small (i.e., 1 to 2 raw score points), and favored the in-person method, consistent with research which directly compared telephone to in-person administration²³. Interestingly, the differential by administration method was more prominent for Letter Fluency compared to Animal Naming (i.e., 2–6 raw score points favoring in-person samples for Letter Fluency). The trend for lower scores in the REGARDS

sample suggests that telephone-based administration and scoring may be negatively affected by limitations in audio quality inherent to the medium.

Administration violations (examiner- and participant-based) were comparable across both fluency measures which indicate there was nothing inherent in the tasks that elicited abnormal behavior in either party. It is possible the disadvantage of an audio-only administration is moderated by the Animal Naming test because the response options are easier to comprehend and express due the overlearned and distinct quality of the network of semantic knowledge that is being probed relative to Letter fluency. In effect, the range of response options is more clear and constrained and may overcome (to an extent) any physical limitations related to the audio-only interface.

Our method of extracting matched samples from REGARDS to compare to the existing published normative and control studies was somewhat limited by the lack of specificity in descriptions of many of the published samples. For example, some studies reported only an education range rather than a mean and SD. If the mean education in such a study were substantially higher than that of REGARDS participants, lower scores achieved in REGARDS would be expected. Likewise, some studies reported only how many participants had attained less than or equal to a high school education vs. greater than a high school education, making it difficult to more precisely match REGARDS participants to these samples. Still others reported only the minimum age cutoff for their sample, such as greater than age 50, prohibiting a precise degree of matching. While these limitations may have contributed to apparent score differences in some instances, it also is possible the REGARDS national U.S. sample captures the true, somewhat broader, range of ability in the given age-education bands. Note that REGARDS cell sizes range from 1,445 to 15,235 participants per cell compared to much smaller participants per cell, generally <100, in the control/comparison group subsamples drawn from published in-person administrations.

Our data reinforce the importance of accounting for mode of administration when interpreting scores in a clinical context. This is especially true for Letter Fluency. Our large sample sizes provide confidence that the levels of performance reflected in the various gender, age, and education groupings are stable and reflect population trends.

Our work extends validation to an epidemiological sample using survey research staff and computer-assisted technology for test administration. We anticipated some discrepancies between scores in our study compared to previous in-person research due to differences in sizes and sources of control/normative samples relative to the REGARDS cohort (e.g., small samples vs. our very large sample; clinical or volunteer samples vs. our population-based cohort; local samples vs. our national cohort). These unmeasured sample differences can introduce known variations in scores. Even so, the scores obtained in REGARDS are comparable to in-person performances across a range of prior studies, and the diversity of the cohort with respect to age, race, and geographic region is a substantial strength for broad generalizability to black and white adults in the US.

Telephone-based, computer-assisted administration and scoring used in REGARDS facilitated the collection of information on administration violations that could be tracked

and used to improve fidelity of data capture including tracking interviewers' use of disallowed prompts. Assistance by someone in the participant's environment was also captured during scoring of audio files. Overall, those who received inappropriate assistance were older and had less education. It is possible that this phenomenon reflects a natural tendency to ease the discomfort of a person who is struggling to produce responses during cognitive testing. The methodology of REGARDS allowed the project to document these actions in a way that traditional in-person assessment has not.

The computer-assisted interface was also subject to technology failures that are not seen in in-person assessments (e.g., equipment failure, technology-based operator error, and inaudibility). In the present study, the occurrence of interviewer-related errors was reduced appreciably over time following corrective actions, as was the rate of equipment failures.

As technology advances, the possibility of using internet-based telephone or video-conferencing presents an intriguing alternative to in-person neuropsychological assessment—methods that can be useful clinically for patients who are homebound or otherwise unable to attend clinic visits. This approach is also advantageous for research in which patient populations are sparse or scattered, rendering in-person clinical evaluations across locations infeasible, excessively costly, or difficult to standardize. Continued dissemination of information on procedures and results from telephone-based projects will be helpful in more fully expanding and exploiting the benefits of neuropsychological assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding Details: This research project is supported by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Service. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or the National Institutes of Health. Representatives of the funding agency have been involved in the review of the manuscript but not directly involved in the collection, management, analysis or interpretation of the data. The authors have no further financial disclosures or conflicts. The authors thank the other investigators, the staff, and the participants of the REGARDS study for their valuable contributions. A full list of participating REGARDS investigators and institutions can be found at <http://www.regardsstudy.org>.

References

1. Castanho TC, Amorim L, Zihl J, Palha JA, Sousa N, Santos NC. Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: a review of validated instruments. *Front Aging Neurosci.* 2014;6. doi:10.3389/fnagi.2014.00016
2. Herr M, Ankri J. A Critical Review of the Use of Telephone Tests to Identify Cognitive Impairment in Epidemiology and Clinical Research. *J Telemed Telecare.* 2013;19(1):45–54. doi: 10.1177/1357633X12474962 [PubMed: 23390209]
3. Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design. *Neuroepidemiology.* 2005;25(3):135–143. doi: 10.1159/000086678 [PubMed: 15990444]
4. Howard VJ, Woolson RF, Egan BM, et al. Prevalence of hypertension by duration and age at exposure to the stroke belt. *J Am Soc Hypertens.* 2010;4(1):32–41. doi:10.1016/j.jash.2010.02.001 [PubMed: 20374949]

5. Validity of the Telephone Interview for Cognitive Status (TICS) in post-stroke subjects - Barber - 2004 - International Journal of Geriatric Psychiatry - Wiley Online Library. <http://onlinelibrary.wiley.com/doi/10.1002/gps.1041/abstract>. Accessed December 5, 2017.
6. Brady CB, Spiro A, McGlinchey-Berroth R, Milberg W, Gaziano JM. Stroke Risk Predicts Verbal Fluency Decline in Healthy Older Men Evidence From the Normative Aging Study. *J Gerontol Ser B*. 2001;56(6):P340–P346. doi:10.1093/geronb/56.6.P340
7. Rafnsson SB, Deary IJ, Smith FB, Whiteman MC, Fowkes FGR. Cardiovascular Diseases and Decline in Cognitive Function in an Elderly Community Population: The Edinburgh Artery Study. *Psychosom Med*. 2007;69(5):425. doi:10.1097/psy.0b013e318068fce4 [PubMed: 17556639]
8. Strauss E, Sherman E, Spreen O. A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Third Edition Oxford, New York: Oxford University Press; 2006.
9. Jones WJ, Williams LS, Meschia JF. Validating the questionnaire for Verifying Stroke-Free Status (QVSFS) by neurological history and examination. *Stroke*. 2001;32(10):2232–2236. [PubMed: 11588306]
10. Soliman EZ, Howard G, Prineas RJ, McClure LA, Howard VJ. Calculating Cornell voltage from nonstandard chest electrode recording site in the Reasons for Geographic And Racial Differences in Stroke study. *J Electrocardiol*. 2010;43(3):209–214. doi:10.1016/j.jelectrocard.2009.10.002 [PubMed: 20004413]
11. Menke A, Casagrande S, Geiss L, Cowie CC. Prevalence of and Trends in Diabetes Among Adults in the United States, 1988–2012. *JAMA*. 2015;314(10):1021–1029. doi:10.1001/jama.2015.10029 [PubMed: 26348752]
12. Olives C, Myerson R, Mokdad AH, Murray CJL, Lim SS. Prevalence, Awareness, Treatment, and Control of Hypertension in United States Counties, 2001–2009. *PLOS ONE*. 2013;8(4):e60308. doi:10.1371/journal.pone.0060308 [PubMed: 23577099]
13. Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK. Norms for Letter and Category Fluency: Demographic Corrections for Age, Education, and Ethnicity. *Assessment*. 1999;6(2):147–178. doi:10.1177/107319119900600204 [PubMed: 10335019]
14. Tombaugh TN, Kozak J, Rees L. Normative Data Stratified by Age and Education for Two Measures of Verbal Fluency: FAS and Animal Naming. *Arch Clin Neuropsychol*. 1999;14(2):167–177. doi:10.1016/S0887-6177(97)00095-4 [PubMed: 14590600]
15. Rankin MW, Clemons TE, McBee WL, Group TA-REDS (AREDS) R. Correlation Analysis of the In-clinic and Telephone Batteries from the AREDS Cognitive Function Ancillary Study. AREDS Report No. 15. *Ophthalmic Epidemiol*. 2005;12(4):271–277. doi:10.1080/09286580591003815 [PubMed: 16033748]
16. Acevedo A, Loewenstein DA, Barker WW, et al. Category fluency test: Normative data for English- and Spanish-speaking elderly. *J Int Neuropsychol Soc*. 2000;6(7):760–769. doi:10.1017/S1355617700677032 [PubMed: 11105466]
17. Crossley M, D'arcy C, Rawson NSB. Letter and category fluency in community-dwelling canadian seniors: A comparison of normal participants to those with dementia of the Alzheimer or vascular type. *J Clin Exp Neuropsychol*. 1997;19(1):52–62. doi:10.1080/01688639708403836 [PubMed: 9071641]
18. Canning SJD, Leach L, Stuss D, Ngo L, Black SE. Diagnostic utility of abbreviated fluency measures in Alzheimer disease and vascular dementia. *Neurology*. 2004;62(4):556–562. doi:10.1212/WNL.62.4.556 [PubMed: 14981170]
19. Diaz M, Kuslansky G. Category size effects in semantic and letter fluency in Alzheimer's patients. *Brain Lang*. 2004;89(1):108–114. doi:10.1016/S0093-934X(03)00307-9 [PubMed: 15010242]
20. Brandt J, Manning KJ. Patterns of word-list generation in mild cognitive impairment and alzheimer's disease. *Clin Neuropsychol*. 2009;23(5):870–879. doi:10.1080/13854040802585063 [PubMed: 19301196]
21. Ravdin LD, Katzen HL, Agrawal P, Relkin NR. Letter and Semantic Fluency in Older Adults: Effects of Mild Depressive Symptoms and Age-Stratified Normative Data. *Clin Neuropsychol*. 2003;17(2):195–202. doi:10.1076/clin.17.2.195.16500 [PubMed: 13680426]

22. Kozora E, Cullum CM. Generative naming in normal aging: Total output and qualitative changes using phonemic and semantic constraints. *Clin Neuropsychol.* 1995;9(4):313–320. doi: 10.1080/13854049508400495
23. Validation of a Cognitive Assessment Battery Administered over the Telephone - Rapp - 2012 - Journal of the American Geriatrics Society - Wiley Online Library. <http://onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.2012.04111.x/abstract>. Accessed December 5, 2017.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

KEY POINTS:

1. Data are from a large, national community-based cohort (REGARDS) from over half of the counties in the contiguous US, including 37–39% African American participants.
2. Methods described in this paper provide clear administration and scoring instructions for telephone-administered verbal fluency measures, Animal Naming and Letter “F” Fluency. These may serve as a model for future longitudinal and epidemiological studies employing telephone-based cognitive assessment.
3. Scores for telephone-administered measures of verbal fluency were comparable to those from other published studies of in person administration methods.
4. Data presented herein, stratified by age, education and gender, can be used as a normative reference.

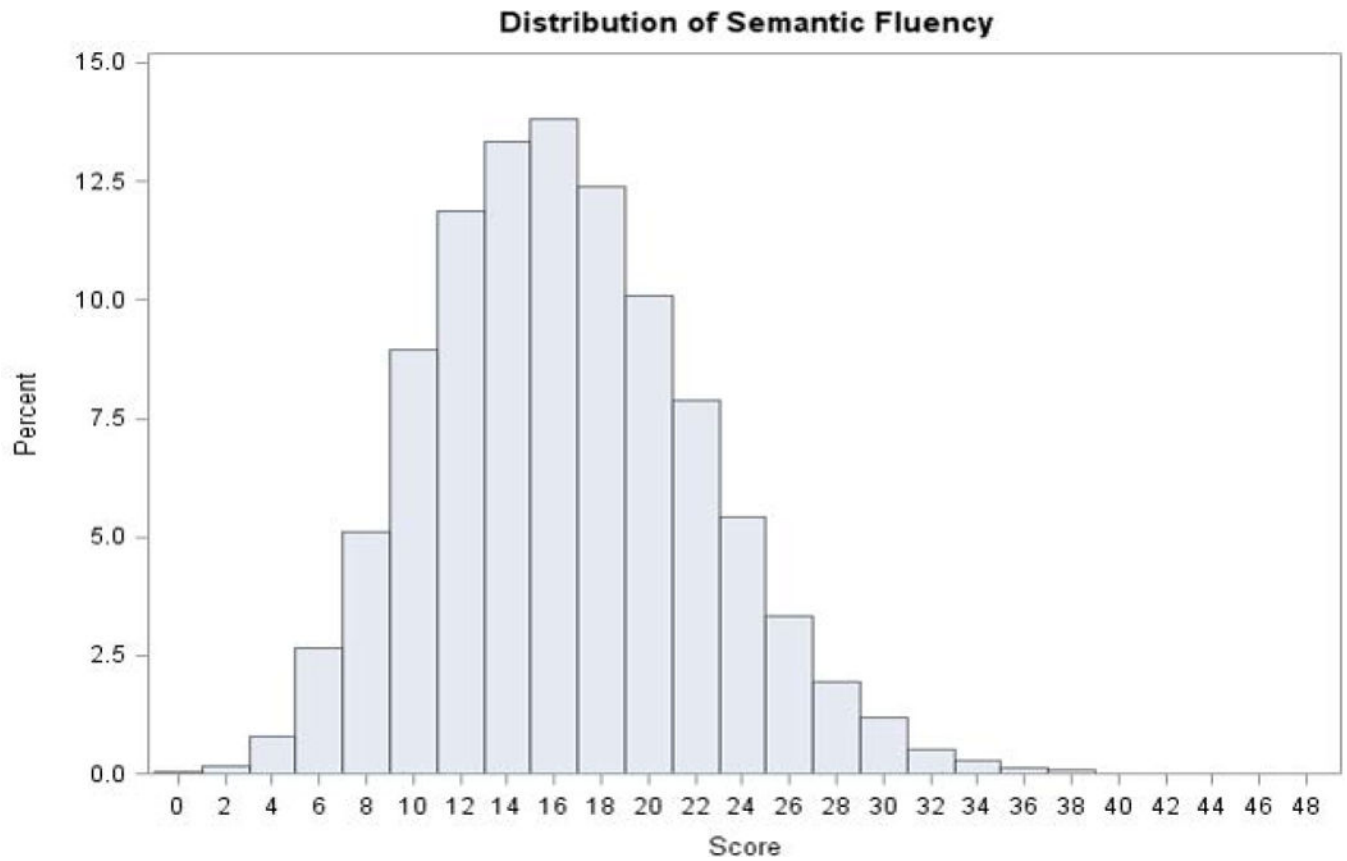


Figure 1a.

Frequency distributions for Animal (Semantic) fluency ($n = 18,072$) in REGARDS. Note: Scores are from participants' first animal fluency assessment. Intrusions and repetitions have been removed.

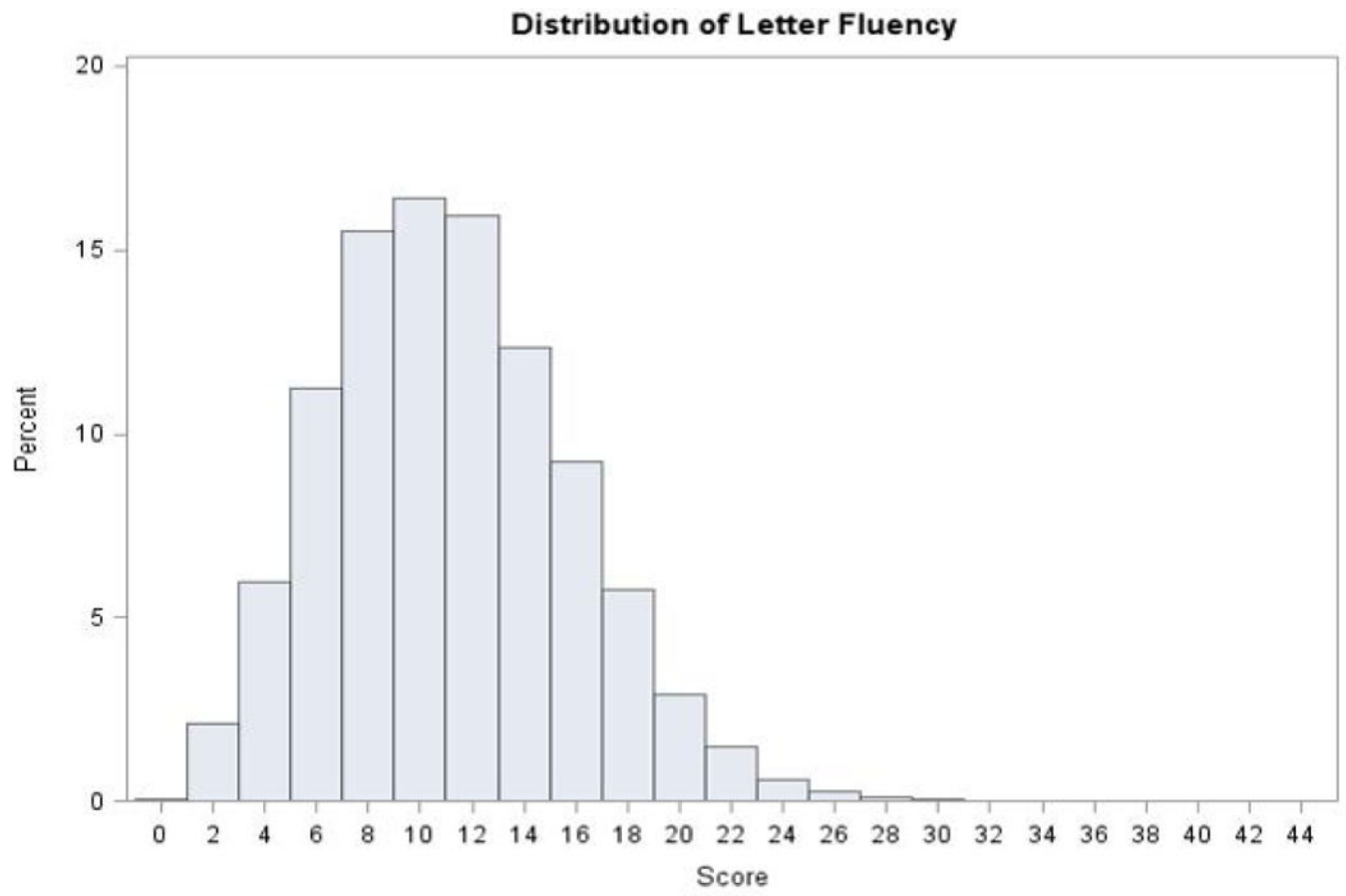


Figure 1b.

Frequency distributions for Letter F fluency ($n = 18,505$) in REGARDS. Note: Scores are from participants' first Letter F assessment. Intrusions and repetitions have been removed.

Table 1.

Sample Characteristics.

	Animal Naming		Letter “F” Fluency	
Gender	<i>N</i>	(%)	<i>N</i>	(%)
Female	10204	(56)	10639	(57)
Male	7868	(44)	7866	(43)
Race				
African American	6976	(39)	6877	(37)
White	11096	(61)	11628	(63)
Age	67.8	(9.2)	69.1	(9.1)
45 - 54 yrs	1498	(8)	1204	(7)
55 - 64 yrs	5789	(32)	5091	(28)
64 - 74 yrs	6785	(38)	7383	(40)
75 - 84 yrs	3368	(19)	4003	(22)
85+ yrs	632	(4)	824	(4)
Education				
0 - 8 yrs	591	(3)	498	(3)
9 - 11 yrs	1275	(7)	1163	(6)
12 yrs/GED	4590	(25)	4633	(25)
13 - 15 yrs	4833	(27)	5007	(27)
16 yrs/college graduate	3701	(20)	3864	(21)
17+ yrs/post graduate	3076	(17)	3336	(18)
Region of Residence				
Non-Stroke Belt	7932	(44)	8024	(43)
Stroke Belt	10140	(56)	10481	(57)
Health				
Hypertension (missing 42/42)	10189	(57)	10305	(56)
Diabetes (missing 621/591)	3396	(19)	3276	(18)
Heart disease (missing 299/292)	2740	(15)	2645	(15)
Left ventricular hypertrophy (missing 5155/4619)	544	(4)	552	(4)
Atrial fibrillation (missing 361/362)	1368	(8)	1351	(7)

Table 2.
Descriptive Statistics for Animal Naming and Letter “F” Fluency by Demographic Subgroups

Variable	Animal Naming			Letter “F” Fluency				
	n (%)	M	SD	Range	n (%)	M	SD	Range
Education (years)								
0 - 8	591 (3)	12.4	4.8	2 - 31	498 (3)	6.7	4.4	0 - 44
9 - 11	1275 (7)	13.0	4.6	3 - 30	1163 (6)	8.0	3.9	0 - 24
12	4590 (25)	15.2	5.1	2 - 41	4633 (25)	9.6	4.2	1 - 29
13 - 15	4833 (27)	16.8	5.3	0 - 41	5007 (27)	10.9	4.4	0 - 30
16	3701 (20)	18.6	5.8	3 - 41	3864 (21)	12.0	4.5	1 - 32
17+	3076 (17)	19.6	6.0	0 - 48	3336 (18)	13.1	4.6	1 - 33
Age (years)								
45 - 54	1498 (8)	20.0	6.2	4 - 48	1204 (7)	12.4	4.6	0 - 32
55 - 64	5789 (32)	18.4	5.7	3 - 43	5091 (28)	11.8	4.6	0 - 32
64 - 74	6785 (38)	16.4	5.5	0 - 41	7383 (40)	10.7	4.6	0 - 44
75 - 84	3368 (19)	14.4	5.2	0 - 36	4003 (22)	9.9	4.4	0 - 27
85+	632 (4)	13.0	4.6	2 - 31	824 (4)	9.6	4.5	1 - 29
Gender								
Female	10204 (56)	16.4	5.8	0 - 48	10639 (57)	11.0	4.7	0 - 44
Male	7868 (44)	17.4	5.8	0 - 43	7866 (43)	10.7	4.6	0 - 33
Total Sample	18,072	16.8	5.8	0 - 48	18,505	10.9	4.7	0 - 44

Comparison to published studies of verbal fluency gathered by in-person assessment relative to demographically matched subsamples from REGARDS

Table 3.

Animal Naming Studies	Control Group from Study			REGARDS Matched-Sample		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Gladys et al. (1999)[†] 55% White, 45% African American						
Age group 50+ years	-	15.28	3.80	1866	12.8	4.7
Education (years)	0 - 11					
	12 - 15	18.05	4.81	9423	16.0	5.3
	16+	19.35	4.42	6777	19.1	5.9
Tombaugh et al. (1999) Canadian Sample; 55.6% Female						
Age (years)	50 - 59	20.10	4.90	3277	19.1	5.9
	60 - 69	17.60	4.70	7272	17.5	5.7
	70 - 79	16.10	4.00	5243	15.3	5.3
	80 - 89	14.30	3.90	1721	13.8	4.9
	90 - 95	13.00	3.80	128	12.3	4.6
Age: 60 - 79 years						
Education (years)	0 - 8	14.40	3.40	434	12.6	4.9
	9 - 12	16.40	4.30	4197	14.6	5.0
	13 - 21	18.20	4.20	7879	17.9	5.5
Age: 80 - 95 years						
Education (years)	0 - 8	13.10	3.80	123	11.1	4.2
	9 - 12	13.90	3.40	632	12.4	4.3
	13 - 21	16.30	4.30	1093	14.8	5.0
Rankin et al. (2005) 97% White; 57% Female Mean Age (<i>SD</i>) = 74.9 (5.0)						
Age (years)	55 - 80	17.60	4.90	14719	16.9	5.7
Acevedo et al. (2000) English-speaking 74% Female						
Total Sample	-	16.20	4.50	18702	16.8	5.8
Age (years)	50 - 59	18.40	4.90	3277	19.1	5.9
	60 - 69	17.10	4.20	7272	17.5	5.7
	70 - 79	15.20	4.30	4874	15.4	5.3

Animal Naming	Control Group from Study				REGARDS Matched-Sample			
Studies	<i>n</i>	<i>M</i>	<i>SD</i>		<i>n</i>	<i>M</i>	<i>SD</i>	
Crossley et al. (1997) Canadian Sample 59% Female Mean Age (<i>SD</i>) = 78.8 (6.8) Mean Educ. (<i>SD</i>) = 9.4 (4.1)	Education (years) [‡]	8 – 12	112	15.00	4.30	4981	15.0	5.1
		13 – 16	154	16.30	4.00	7314	17.9	5.5
		17+	5	18.80	5.40	2688	20.0	5.8
	Gender	Male	82	16.20	4.60	6681	17.8	5.7
		Female	234	16.30	4.50	8742	16.7	5.7
	Age (years)	65 – 74	144	14.20	4.30	6785	16.4	5.5
		75 – 84	343	14.20	3.80	3368	14.4	5.2
		85	148	12.50	3.80	632	13.0	4.6
	Education (years) [§]	0 – 6	149	12.10	3.10	489	12.0	4.7
		7 – 9	169	13.40	3.80	-	-	-
Letter “F” Fluency Studies Canning et al. (2004) Mean Age (<i>SD</i>) = 70.10 (6.7) Mean Education (<i>SD</i>) = 16.0 (2.9) Diaz et al. (2004) 67% Female Mean Age = 79.13 Mean Education = 11.33 Brandt & Manning (2009) 60% Female Mean Age (<i>SD</i>) = 74.57 (5.73) Mean Education (<i>SD</i>) = 15.65 (2.64) Ravdin, et al. (2003) 78.5% Female		10 – 12	203	14.10	3.90	3744	13.9	4.9
		13	113	16.30	4.10	6547	16.8	5.5
	Gender	Male	258	14.20	4.20	5025	16.3	5.6
		Female	377	13.60	3.90	5760	14.9	5.3

Animal Naming	Studies	Control Group from Study			REGARDS Matched-Sample		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Mean age (<i>SD</i>) = 74.82 (6.63)		80	15.50	4.1	5768	10.4	4.5
Mean Educ. (<i>SD</i>) = 15.57 (2.67)		35	15.43	4.34	1445	9.9	4.3
Kozora & Cullum (1995)	Age (years)	70 – 79					
		80 – 85					
		50 – 59	41	14.05	2576	12.2	4.6
		60 – 69	43	15.69	7234	11.4	4.6
		70 – 79	47	15.98	5768	10.4	4.5
		80 – 89	43	14.21	2040	9.9	4.3

Note. Mean age and education are provided if comparison sample is not stratified on these variables.

[†]Only participants in the 50+ age group are included in this table; *n* unavailable for stratified groups.

[‡]REGARDS education categories: 9–12 years, 13–16 years, 17+ years.

[§]REGARDS education categories: 0–8 years, 9–12 years, 13+ years.

Table 4.

Administration violations (*n*, %) by type and task.

Violation	Animal Naming (<i>n</i> = 25,478)	Letter “F” Fluency (<i>n</i> = 22,250)
Inappropriate interviewer prompt	687 (2.7)	510 (2.3)
Help from person in participant’s environment	53 (0.20)	57 (0.25)
Invalid files, < 50 sec allowed for response	313 (1.2)	421 (2)
Technical difficulties, no audio	5167 (20)	1169 (5)

Note. Files with Prompts and/or Help were retained for potential inclusion in analyses prior to removal of cases meeting additional exclusion criteria. Invalid files and those with Technical Difficulties/ No Audio were not used in analyses.